# Report on the First International Workshop on Energy Data Management (EnDM 2012)

Torben Bach Pedersen
Center for Data-intensive Systems (Daisy)
Aalborg University
9220 Aalborg, Denmark
tbp@cs.aau.dk

Wolfgang Lehner
Database Technology Group,
Dresden University of Technology
01062 Dresden, Germany
{firstname.lastname}@tu-dresden.de

Gregor Hackenbroich
Data Management and Analytics
SAP Research, SAP AG
01187 Dresden, Germany
gregor.hackenbroich@sap.com

## 1. INTRODUCTION

The energy sector is one of the most active application domains being forced to re-think the current practice and apply data-management based IT solutions to provide a scalable and sustainable supply and distribution of energy. Challenges range from energy production by seamlessly incorporating renewable energy resources over energy distribution and monitoring to controlling energy consumption. Decisions are based on huge amounts of empirically collected data from smart meters, new energy sources (increasingly RES - renewable energy sources such as wind, solar, hydro, thermal, etc), new distributions mechanisms (Smart Grid), and new types of consumers and devices, e.g., electric cars.

Energy is at the top of the worldwide political agenda, e.g., due to global warming concerns and recent nuclear accidents. Ambitious goals for reductions of energy consumption and CO2 emissions have been formulated, e.g., the EU 20-20-20 goals (20% renewable energy, 20% better energy efficiency, and 20% CO2 reduction by 2020), with much more ambitious goals set for 2030 and 2050. This situation is reflected by increasing attention in research funding schemes such as the EU 7th Framework program as well as national programs. A recent trend in these programs is joint calls involving both energy and IT partners. Data management is at the heart of this development, as witnessed by the following story headlines from key players: "The Smart Grid Data Deluge" (O'Reilly Radar); "Big data for the Smart Grid" (theenergycollective); "The Coming Smart Grid Data Surge" (SmartGridNews.com).

There is thus a need for focusing on data management within the energy domain. The International Workshop on Energy Data Management (EnDM) focuses on conceptual and system architecture issues related to the management of very large-scale data sets specifically in the context of the energy domain. The overall goal of the EmDM workshop is to bridge the gap between domain experts and data management scientists on the one hand. Additionally, the workshop's goal is to create awareness of this upcoming and very challenging application area. For the workshop's research program, the organizers solicited contributions that push the envelope towards novel schemes for large-scale data processing with special focus on energy data management.

The First International Workshop on Energy Data Management (EnDM'12)[1] was held in conjunction with EDBT 2012 in Berlin, Germany on March 30, 2012. This half-day event brought together researchers and engineers from academia and industry to discuss and exchange ideas related to energy data management and related topics. The workshop featured one industrial keynote, five research papers, and finished off with a discussion. The accepted papers spanned a number of exciting topics within energy data management, including (in no particular order) smart grid architectures, smart grid specific data management challenges, and the use of gamification in active demand response, as well as related issues such as energy efficient file access and energy environmental impact data management. The proceedings of the workshop was published in a joint volume of all EDBT/ICDT 2012 workshops [1].

## 2. INDUSTRIAL KEYNOTE

The keynote was given by Dr. Kevin Brown, Chief Architect for Informix Dynamic Server at IBM, and was entitled "*The Massive Data Challenge - A unique approach to handling smart meter data with a hybrid database*". The talk first outlined the massive challenges related to efficient management of very large amounts of electricity meter data. It then went on to describe the Informix Benchmark for Meter Data Management including the specifics of the captured data and the query workload to process over it. The talk introduced a specific instance of this benchmark, the so-called "100 Million Meter Benchmark" [2], which simulates 100 million meters being read every 15 minutes. The talk also delved into the specifics of the Informix TimeSeries extension, including its optimized physical data storage and loading strategies, as well discussing how it could be used to handle the Meter Data Benchmark much more effieciently than traditional RDBMSes. The solution was shown to handle both the benchmark and seveal real-world cases from US energy companines much more efficiently than traditional solutions. In addition to providing this exciting technical contribution, Dr. Brown provided his valuable industry perspective on the remaining papers in the workshop.

[1] http://endm2012.endm.org

## 3. RESEARCH PAPERS

The paper by Masaru Iritani and Haruo Yokota entitled "*Effects on performance and energy reduction by file relocation based on file-access correlations*" considered the energy-efficient placement and relocation of files across a set of distributed hard disk drives (HDDs) with the goal of reducing the energy consumption of the drives while keeping the access performance of the system at the same level. Previous approaches have mainly located frequently accessed files together on a few drives in order to enable spin-down of the remaining drives, but this causes significant energy consumption for spin-up of these drives when accessing infrequently accessed data, especially when some files that tend to be used together are placed on many different drives. The paper goes further by proposing a novel method called PLECO (Placement of files for Latency and Energy Consumption Optimization). This method tries to locate correlated files on the same drive, and thus both reduce power consumption further while also improving the system performance. The simulated evaluation of PLECO indicates that it can reduce both the energy consumption and the access latency by up to 32% and 92%, respectively, compared with a baseline system.

The next paper was by Benjamin Bertin, Vasile-Marian Scuturici, Emmanuel Risler, and Jean-Marie Pinon and was entitled "*A semantic approach to life cycle assessment applied on energy environmental impact data management.*" The paper concerned semantic web-based modeling of lifecycle assessment for energy environmental impact. Specifically, the paper focused on the life cycle inventory stage of life cycle assessment, which decomposes a life cycle into its individual economic activities. Modeling this is complex due to the large amounts of elementary processes and interdependency links. The paper proposes a semantic approach for the modelling of life cycle inventory databases which in comparison with earlier work offers a more comprehensible model. The model is explained and illustrated with life cycle inventory data for the U.S. electricity production.

The paper by Matthias Boehm, Lars Dannecker, Andreas Doms, Erik Dovgan, Bogdan Filipic, Ulrike Fischer, Wolfgang Lehner, Torben Bach Pedersen, Yoann Pitarch, Laurynas Siksnys, and Tea Tusar called "*Data management in the MIRABEL smart grid system*" focused on the data management challenges of a specific approach to the smart grid. The motivation for the paper is that Renewable Energy Sources (RES) are becoming increasingly important to reduce greenhouse gas emissions and will take up a much larger share of the energy production. This leads to a number of challenges such as balancing energy supply and demand since RES cannot be scheduled. The paper addresses the balancing challenge by specifically presenting the MIRABEL project and its Energy Data Management System (EDMS) which uses the flexibilities available in the electricity demand, e.g., dishwasher, electric vehicles, etc., to efficiently balance energy demand and supply. The major novel concept of MIRABEL are so-called *flex-offers* that explicitly capture intended energy use and the flexibilities in time, amount, and price that are associated with it. A MIRABEL-based EDMS will eventually consist of millions of heterogeneous nodes, each incorporating a number of advanced components such as flex-offer aggregation, forecasting, scheduling, and negotiation. The paper describes each of these components and their interaction while focusing on the data managent challenges that arise. The challenges include effective aggregation of flexibilities, tight integration of forecasting, both for functionality and forecasted data, the interplay between aggregation, forecasting, and scheduling, and the monetization of the flexibilities. The experimental results show that the proposed EDMS is indeed feasible.

The next paper by Benjamin Gnauk, Lars Dannecker, and Martin Hahmann, entitled "*Leveraging gamification in demand dispatch systems*," focused on how to involve energy consumers more actively in so-called demand-side management techniques to help optimize the grid's efficiency and a better utilize renewable energy sources. The paper focuses on so-called demand dispatch systems, where consumers must proactively communicate their flexibilities. A standard incentive is monetary compensation, but this is often not enough to motivate the individual consumer for a sustainable participation. The approach proposed by the paper instead uses *gamification* as a motivational framework. Here, well-known game mechanics instruments, e.g., point awards and leaderboards, are used to engage the consumers. The paper explains the special scoring system used and how it is combined with aspects of social competition in a user interface that helps consumer define and management their flexible energy demands. The paper reports on an initial user study which shows that the user acceptance is high and that the system can potentially engage many consumers.

Finally, the paper by Daniel Rech and Andreas Harth called "*Towards a decentralised hierarchical architecture for smart grids*" considered the technical architecture for smart grids. Specifically, the paper presented a hierarchical distributed communication and control architecture for Smart Grids. The topology of the proposed architecture accomodates the decentralised nature and large sizes of smart grid systems by having multiple layers in order that ensures both robust and flexible data access and resource allocation. The paper describes a specific use scenario with a number of different smart grid actors, and further develops an architecture for this scenario based on the Linked Data principles known from the semantic web area. Further, the authors propose a simple language that can express allocation constraints. They also map the resource allocation problem into a constraint satisfaction problem. The paper finally provides initial experimental results within the tasks of decentralised data access and resource allocation for smart grids.

## 4. DISCUSSION AND OUTLOOK

At the end of the workshop, a lively discussion took place among the participants, the conclusion of which are included in this section along with some post-workshop reflections.

If we first look at the topics of the presented papers, we note that they span a very wide range of topics, ranging from low-level technical issues within data management and communication, over conceptual level modeling, to the integration of user interaction aspects. The papers also cover both energy systems and the energy consumption of IT systems themselves. This is a reflection of the fact that the journey smart grid is long and requires tight collaboration between many different areas not just within computer science itself, but also including inter-disciplinary collaborations with other sciences.

Next, when looking at the topics which occcured in the Call for Papers, but not within the accepted (or submitted) papers, we see that topics such as data security and privacy and data mining techniques for energy data are missing. We believe this is not because the topics are not important, but rather due to the fact that energy data management is still new, and more pressing issues must be solved before considering such topics. While most papers are based on small case studies, there were no papers describing large industrial case studies of already running systems. We again attribute this to the fact that smart grids are still in development.

The workshop discussions identified a number of issues that must be resolved in order to better unify and leverage the many concurrent research activities within energy data management. The first such issue was the lack of common definitions of data and information concepts within the area, e.g., community-wide agreed-upon standard ontologies specifying common concepts. Another issue was the lack of standardization of the units of the technical architecture within smart grid systems, e.g., which types of layers exist, and what the nodes at each layer does. Such standards already exist at the business level of the energy sector, e.g., for standardizing the different types of actors in smart grid setups, but not yet at the more technical levels.

As the final words, we can safely conclude that there is a large demand for further work in the area of energy data management, including a need for venues that focus on this issue. The EnDM workshop series will continue at EDBT 2013 in Genoa where the 2nd International Workshop on Energy Data Management will be held on March 22, 2013[2]. For the 2nd edition of the workshop, it is the intention for organizing a special issue of a journal for extended versions of the best papers.

## 5. ACKNOWLEDGEMENTS

The EnDM Chairs would like to thank all the authors of submitted papers for their interest in the workshop and the high quality of the submitted papers. We would also like to thank the distinguished PC members for their careful and dedicated work, both during the reviewing and the discussion phases. Finally, we would like to thank the Organizing Committee of EBDT/ICDT 2012, especially the Workshop Chair Divesh Shrivastava and the Proceedings Chair Ismail Ari. for their support in organizing the EnDM 2012 workshop.

[2]http://endm2013.endm.org

## 6. REFERENCES

[1] D. Srivastava and I. Ari, editors. *Proceedings of the 2012 Joint EDBT/ICDT Workshops, Berlin, Germany, March 30, 2012*. ACM, 2012.

[2] Unprecedented performance and scalability demonstrated for meter data management. Technical report, IBM Corporation, Online at http://www.ibm.com/developerworks/forums/thread.jspa?messageID=14685713, 2011.